The influence of receptor positioning on chemotactic information

H. Nguyen, P. Dayan, G.J. Goodhill

Queensland Brain Institute, The University of Queensland, St. Lucia, QLD 4072, Australia
School of Mathematics and Physics, The University of Queensland, St. Lucia, QLD 4072, Australia
Gatsby Computational Neuroscience Unit, UCL, London, UK

ABSTRACT

Chemotaxis, or gradient following, is important in many biological systems, but suffers from noise. How receptors are positioned on the cell or sensing device influences the quality of the inferences they can support about the gradient, suggesting that their configuration might be optimised. We show that for an elliptical sensing device, inhomogeneous receptor placement could be a potential approach for cells to eliminate bias in the posterior distribution of the gradient direction. We use information theory to calculate the mutual information between the gradient and the binding pattern, thus finding the optimal receptor arrangement for gradient sensing.

1. Introduction

Many biological systems rely on chemotaxis. These include neutrophils migrating to sites of inflammation (Downey, 1994), the slime mold Dictyostelium discoideum hunting for food (Swaney et al., 2010), and neuronal growth cones navigating to find their targets in the developing nervous system (Mortimer et al., 2008; Lowery and Van Vactor, 2009). The ability of such sensing devices to detect chemical gradients depends sensitively on unavoidable stochastic fluctuations due to the limited numbers of receptors, intracellular signalling molecules, and ligand molecules available in the gradient itself (Berg and Purcell, 1977; Bialek and Setayeshgar, 2005). Detecting a gradient can thus be seen as a paradigmatic problem of reasoning in the face of uncertainty (Mortimer et al., 2009). Here we focus on noise due to receptor binding fluctuations.

A powerful approach for analysing such problems is to consider the optimal statistical inference that an ideal observer would perform (Andrews and Iglesias, 2007; Mortimer et al., 2009; Fuller et al., 2010; Hu et al., 2010, 2011a; Mortimer et al., 2011). This involves combining available information with prior assumptions. However a critical unanswered question is the extent to which some spatial distributions of receptors admit better gradient detection than others. Starting from the familiar model of the sensing device (hereafter ‘cell’) as a two-dimensional ellipse with receptors distributed on the surface, we derive the mutual information between the gradient and binding pattern as a target quantity to maximise in order to achieve optimal inference.

A recent theoretical analysis shows that with a uniformly distributed set of receptors, an elliptical cell can make incorrect inferences about the gradient when the concentration and the gradient steepness are low (Baba et al., 2012). Surprisingly, the cell has a strong bias to infer that the gradient is parallel to the minor axis, regardless of the actual gradient direction. This is because equal spacing of receptors on a non-circular surface leads to highly unequal variances in the estimates of the x and y components of the gradient. Here we show that this can be overcome by a nonuniform placement of receptors so that the inference is free of biases due to the shape of the cell.

2. Model

We consider the cell as estimating the gradient $\hat{\mu}$ of a spatial function $C(\vec{r}) = C_0 \exp(\mu \cdot \vec{r})$. Receptor positions $\vec{r}_i$ relative to the ‘standard’ length scale 10 µm and the gradient $\hat{\mu}$ is dimensionless. We assume that the information available about $C$ consists of independent binary random variables $b_i$ representing the bound and unbound states of a set of $n$ receptors located at positions $\vec{r}_i \in \mathbb{R}^2$, $i = 1 \ldots n$. Standard Michaelis–Menten kinetics implies that the binding probability of each receptor is

$$P(b_i = 1) = \frac{C(\vec{r}_i)}{K_d + C(\vec{r}_i)}$$

with $K_d$ being the dissociation constant. The likelihood function of the complete binding state is

$$L_b(\hat{\mu}, C_0) = \prod_{i=1}^{n} \left( \frac{C(\vec{r}_i)}{K_d + C(\vec{r}_i)} \right)^{b_i} \left( \frac{K_d}{K_d + C(\vec{r}_i)} \right)^{1-b_i}$$
where the logarithm is

$$\ln L_b(\mu, C_0) = \frac{n}{2} \ln \left( \frac{C(\bar{r}^T_1)}{K_d} \right) - \frac{n}{2} \ln \left( \frac{K_d + C(\bar{r}^T_1)}{K_d} \right)$$

The cell should combine likelihood information with its a priori estimate of the gradient. The prior has two components: the first is the direction $\phi = \mu$, which is conventionally represented as a Vines distribution as in Hu et al. (2011a):

$$P(\phi) = \frac{\exp(\kappa \cos(\phi - \delta))}{I_0(\kappa)}$$

where $\delta$ is the prior bias of the cell regarding the gradient direction, $\kappa$ is the strength of bias, and $I_0(\kappa)$ is the modified Bessel function of the first kind. This prior could be determined by previous measurements, as in a filtering scheme, or by an intrinsic bias. The second component is the strength $\kappa = |\mu|$ of the gradient. For convenience, we consider a simple, half-Gaussian form for this $P(s) = 2\sqrt{\beta}/\pi H(s)\exp(-\beta s^2)$, where $\beta$ parameterizes the uncertainty. This favors small gradients, a conclusion invited by the exquisite sensitivity of many sensing systems (Mortimer et al., 2009; Mao et al., 2003). However, its precise form is not expected to influence the results very strongly, provided it is smooth and covers the range of relevant values. We consider these two components to be independent, making the overall prior $P(s, \phi) = P(s) \times P(\phi)$.

Expanding the likelihood function to second order around $0$ in $\mu$:

$$\ln L_b \approx \frac{n}{2} \ln \frac{C_0}{K} + \frac{1}{2} \bar{r}^T S^{-1} \bar{r} - n \ln K + n \ln K_0$$

where

$$\Delta_b \bar{r} = \sum_{i=1}^{n} \left( \bar{r}_i - \frac{C_0}{C_0 + K} \right)$$

leads to the maximum likelihood estimate (MLE)

$$\mu^{ML} = S^{-1} \Delta_b \bar{r}$$

This formula is more general than that derived in Hu et al. (2010) since it does not assume a circular cell or a uniform distribution of receptors on the cell's surface. The average binding probability $E[b_i]$ at each receptor is

$$E[b_i] = \frac{C_0 \exp(\mu \cdot \bar{r})}{K + C_0 \exp(\mu \cdot \bar{r})} \approx \frac{C_0}{K + C_0 \sum \exp(\mu \cdot \bar{r})},$$

and therefore

$$E[|\mu|] \approx \frac{C_0 K}{(C_0 + K)^2}$$

confirming that the expectation of $\mu^{ML}$ over all possible binding patterns is the actual gradient.

In the large $n$ limit, the properties of the MLE ensures that $\mu^{ML} \sim N(\hat{\mu}, S^{-1})$. $S^{-1}$ is the covariance matrix of the maximum likelihood estimate and only depends on the positions of the receptors, not the shape of the cell. We call $S$ the ‘receptor matrix’ as it ultimately encodes information about the receptor arrangement. As $S$ is a symmetric matrix it can be diagonalised, implying that there exists a coordinate system defined by the two eigenvectors of $S$ (shown in Fig. 1) such that the two orthogonal components of $\mu^{ML}$ are uncorrelated, and their variances are the eigenvalues of the matrix $S^{-1}$. Henceforth, we will define all angles relative to this coordinate system, with $x, y$ axes identified with the first and second eigenvectors of $S^{-1}$. Note that these axes will in general be different from the axes of the elliptical cell.

3. Eliminating bias

For certain receptor distributions for which $\sigma_1 \neq \sigma_2$, the variances in $\mu_x^{ML}$ and $\mu_y^{ML}$ can differ, causing the cell consistently to estimate the gradient direction

$$\phi = \tan^{-1} \left( \frac{\mu_y^{ML}}{\mu_x^{ML}} \right)$$

parallel to its minor axis at low concentration or gradient steepness, as seen in Baba et al. (2012). At first glance, this result might be counter-intuitive. However, if $\sigma_1 > \sigma_2$, equivalent to a ‘receptor ellipse’ elongated in the $x$ direction, the cell can much more easily detect the asymmetry in the concentration in the $x$ direction (low variance) than in the $y$ direction (high variance). The inequality in variances leads to bias in the MLE due to the highly nonlinear nature of the function $\tan^{-1}$. Therefore, at shallow gradients the estimated direction of the gradient has a tendency to favor the minor axis (the $y$ direction). The estimated direction also has higher variance if the true gradient is in the $x$ direction than if it is in the $y$ direction as illustrated in Fig. 2.

For simplicity we assume that the cell is only interested in the direction $\phi$ and not the magnitude. In order to find the maximum a posteriori (MAP) estimate for the actual gradient direction $\phi_{true}$, we seek to solve $\phi_{MAP} = \arg \max_{\phi} P(\phi|Z)$ where

$$P(\phi|Z) \propto \int P(\phi|Z, s) P(s) P(\phi|s) \, ds$$

$$\propto \frac{1}{\sqrt{A}} \exp \left( \frac{B^2}{4A} - C \right) \left( 1 - \text{erf} \left( \frac{-B}{2\sqrt{A}} \right) \right)$$

where

$$A = 1/2(\sigma_1^2 \cos^2 \phi + \sigma_2^2 \sin^2 \phi + \beta)$$

$$B = Z_1 \cos \phi + Z_2 \sin \phi$$

$$C = K_\phi - K_C$$

Fig. 1. Schematic problem representation. The orange dots represents receptors. The axes of the coordinate system are the two eigenvectors of the ‘receptor ellipse’ matrix

$$S = \sum_i r_i r_i^T C_0 K \left( C_0 + K \right)^2$$

which might or might not coincide with the axes of the actual cell (red). The two axes of the ‘receptor ellipse’ determines the properties of gradient estimation. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

We define $1/\sigma_1^2$ and $1/\sigma_2^2$ to be the corresponding eigenvalues of the matrix $S^{-1}$ and

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \mu_{x}^{ML} \sigma_1^2 \\ \mu_{y}^{ML} \sigma_2^2 \end{bmatrix} = \sum_j \sigma_j^2 b_j (r_i \cos \varphi_j - \mu_j + C_0)$$

where $r_i, \varphi_j$ are the positions of the receptors in polar coordinates, and thus recover the familiar Gaussian approximation for the likelihood function (Hu et al., 2010):
Although weak (column 1). As bias due to the inequality between \( \phi_\text{7} \), for \( \phi_\text{7} \) is a strong bias for the posterior distribution to peak at 0 or case and \( \phi_\text{7} \) strongly favors values close to \( \pm \pi/2 \). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

\[
C = \frac{Z_1^2}{2\sigma_1^2} + \frac{Z_2^2}{2\sigma_2^2} - \kappa \cos(\phi - \delta)
\]

Denoting

\[
\hat{\phi} = \tan^{-1}\left(\frac{Z_1}{Z_2}\right)
\]

as the maximum likelihood estimate of the gradient direction and \( |Z| = \sqrt{Z_1^2 + Z_2^2} \) representing the asymmetry in the receptor occupancy, we numerically calculate \( P(\phi, |Z|) \) for different values of \( |Z| \). Fig. 3 shows the relationship between \( \hat{\phi} \) and the posterior estimate of the gradient direction at three values of \( |Z| = 0.5, 2 \) and \( 7 \), for \( \sigma_1 = 3, \sigma_2 = 7 \) (row 1), which we loosely call the 'elliptical' case and \( \sigma_1 = 5, \sigma_2 = 5 \) (row 2), the 'circular' case. The black lines represent the maximum a posteriori estimates for each case. The prior distribution for all cases is

\[
P(s, \phi) \propto \exp(0.5(\phi + \pi/2) + 2s^2).
\]

For the elliptical case, the posterior is influenced by three factors: the bias due to the inequality between \( \sigma_1 \) and \( \sigma_2 \) represented by the term \( A \) in Eq. 3, the prior represented by the last term of \( C \), and the data captured by the term \( B \). Since \( A \) is minimised when \( \phi = 0 \) or \( \pi \), there is a strong bias for the posterior distribution to peak at 0 or \( \pi \) (aligned with the direction of \( \phi_\text{7} \)), which can overwhelm the prior when \( |Z| \) is weak (column 1). As \( |Z| \) becomes stronger, this tendency remains, and although \( \phi_\text{MAP} \) follows \( \phi \) more closely (column 3), the cell is more precise at estimating gradient directions pointing along the minor axis of the receptor ellipse than along the major axis, consistent with the results in Baba et al. (2012). The posterior is more sharply peaked when \( \phi = 0 \) or \( \pi \) because when \( B \) is maximum, \( A \) is minimum, and vice versa. If \( \phi = \pi/2 \) or \( 3\pi/2 \), the maxima and minima of \( B \) and \( A \) are ‘in phase’, therefore reducing the variance. At lower values of \( |Z| \) (column 1 and 2), there are discontinuities in \( \phi_\text{MAP} \) around \( \phi = \pi/2 \) and \( 3\pi/2 \). The graph in column 2 shows the intermediate case between the two extremes of the estimates being dominated by the shape bias (column 1) and dominated by the binding observation.

In contrast, the circular case does not have the bias due to the shape (row 2). If \( |Z| \) is weak, the gradient estimates are dominated by the prior. When \( |Z| \) is large, the maximum a posteriori estimates become almost equal to the maximum likelihood estimates \( \phi \) (row 2, column 3). The posterior distribution is more sharply peaked when \( \phi \) is near the prior mean \( 3\pi/2 \), meaning that the data agree with the prior knowledge.

The estimation problem can be visualised using numerical simulations. Recalling that \( Z_1, Z_2 \) are Gaussian random variables with means \( \sigma_1^2 \cos \phi_{\text{true}}, \sigma_2^2 \sin \phi_{\text{true}} \) and variances \( \sigma_1^2 \) and \( \sigma_2^2 \) respectively, and arbitrarily setting the prior mean \( \delta = -\pi/2 \), we can visualise the distribution of all MAP estimates for various combinations of parameters (Fig. 4). The red plots are the posterior distribution \( P(\phi, |Z|) \) weighted by the empirical probability density \( P(Z/\sigma_{\text{true}}, \phi_{\text{true}}) \). For \( \sigma_1 \neq \sigma_2 \) (columns 1–4, row 1), when the prior \( \kappa \) is weak and the gradient \( s \) is shallow, the estimates are strongly biased towards the minor axis of the ‘receptor ellipse’ \( S \) (c.f. Baba et al., 2012). The discontinuities of \( \phi_\text{MAP} \) at angles parallel to the major axis of the ‘receptor ellipse’ are also observed, consistent with row 1 of Fig. 2. When the gradient is stronger (row 2), the estimates follow the true gradient more faithfully, and the performance is better when the true gradient is aligned with the minor axis than when it is aligned with the major axis (row 2, column 1 vs 3). Even when the prior is strong and the true gradient is shallow (row 3), the estimates can still be strongly influenced by the cell’s bias towards the minor axis. When the prior and the gradient are both strong (row 4), the estimates represent a compromise between the true gradient, the prior and the minor axis.

Meanwhile, when \( \sigma_1 = \sigma_2 \) (columns 5–6), this bias towards the minor axis is eliminated. Note that it is not the case that there is a certain strength of receptor heterogeneity that is required to overcome the bias imposed by cell shape, rather any receptor arrangement that satisfies \( \sigma_1 = \sigma_2 \) is a solution. The distribution of the estimates then becomes a compromise between the prior directional bias and the measurements. If the actual gradient is shallow and the prior over the direction is weak (row 1), the estimates can fluctuate greatly however, the estimates stay centred around the true gradient direction. If the prior is strong (row 3), the measurements contribute little to the estimates. The estimates become more accurate when the prior is weak and the gradient is steep (row 2).

We now focus on the special case \( \sigma_1 = \sigma_2 = \sigma \), since it avoids the bias coming from the spatial arrangement of the receptors. In the next section, we will show that by arranging receptors in such a way that \( \sigma \) is maximised, the cell maximizes the information about the gradient from its binding patterns. This optimization is subject to the constraint that the receptors have to be on the cell surface. We know that the covariance of \( \mu = \mu_{\text{ML}} \) is the matrix \( S^{-1} \) with eigenvalues \( 1/\sigma_1^2 \) and \( 1/\sigma_2^2 \) from Eq. 2. Imposing \( \sigma_1 = \sigma_2 \) implies that the diagonal terms of \( S \) must be equal:

\[
\sum_{i} C_0K_d \sigma_{i}^2 \cos^2 \frac{\theta_i}{(C_0 + K_d)} = \sum_{j} C_0K_d \sigma_{j}^2 \sin^2 \frac{\theta_j}{(C_0 + K_d)}
\]

and maximising \( \sigma \) requires maximising the eigenvalue of \( S \) or maximising \( S_{1,2} = \sigma_{1,2}^2 - \sigma_{1,2}^2 \). Thus the cross correlation term vanishes: \( \Sigma \sigma_{i}^2 \cos \frac{\theta_i}{\sin \theta_i} = 0 \). Recalling the ellipse equation \( x^2/a^2 + y^2/b^2 = 1 \), we can easily see that

\[
\sigma_1^2 = \frac{nC_0K_d}{(C_0 + K_d)^2} \frac{a^2b^2}{a^2 + b^2}
\]

which leads to

\[
\sigma_1 = \sigma_2 = \frac{nC_0K_d}{(C_0 + K_d)^2} \frac{a^2b^2}{a^2 + b^2}
\]

Three example receptor arrangements that satisfy the above conditions are shown in Fig. 5. We found these numerically; however, extra insight into appropriate arrangements comes from
Fig. 3. Numerical comparison of the posterior distribution of the gradient direction $\phi$ given the maximum likelihood estimate $\hat{\phi}$ for $\sigma_1 = 3, \sigma_2 = 7$ (row 1) and $\sigma_1 = \sigma_2 = 5$ (row 2), using the same prior $P(s, \phi) \propto \exp(0.5 \cos(\phi+\pi/2)) \exp(-2s^2)$. The maximum and minimum in each plot from left to right are (0.06, 0.25), (0.05, 0.34), (0.19, 0.99), (0.08, 0.26), (0.06, 0.30), (0.02, 0.50). The black lines represent the maximum a posteriori estimates. The first row shows a clear nonuniform quality in the posterior distributions given different $\hat{\phi}$. The posterior distribution is sharper when $\hat{\phi}$ aligns with $\sigma_1$ ($\hat{\phi} = 0$ or $\pi$) than when $\hat{\phi}$ aligns with $\sigma_2$ ($\hat{\phi} = \pi/2$ or $3\pi/2$).

<table>
<thead>
<tr>
<th>$(\kappa, s_{true})$</th>
<th>$(\sigma_1, \sigma_2)=(3,7)$</th>
<th>$(\sigma_1, \sigma_2)=(7,3)$</th>
<th>$(\sigma_1, \sigma_2)=(5,5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.1,0.1)</td>
<td>$\phi_{true} = 0$ $\pi/3$</td>
<td>0</td>
<td>$\phi_{true} = 0$ $\pi/3$</td>
</tr>
<tr>
<td>(0.1,0.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,0.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,0.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. The distribution of MAP estimates for different gradient magnitudes and gradient direction priors. The x and y axes are defined by the matrix $S$ as in Fig. 1. As the prior over $s$ plays little role, we assume that the same prior $P(s, \phi) \propto \exp(0.5 \cos(\phi+\pi/2)) \exp(-2s^2)$ and $\delta = -\pi/2$ for all cases. Columns 1–4: $\sigma_1 \neq \sigma_2$. Columns 5–6: $\sigma_1 = \sigma_2$. The red plot is the distribution $P(\phi_{posterior}|s_{true}, \phi_{true})$, arbitrarily scaled for easy visualization. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)
4. Optimising the mutual information

A canonical way to quantify the quality with which the receptors constrain the estimate of the gradient is via the mutual information between the gradient direction estimate and the binding pattern, defined as

\[ I(\phi, Z) = \int \log P(\phi|Z)P(\phi, Z) \, d\phi \, dZ - \int \log P(\phi)P(\phi) \, d\phi \]

The second term is fixed, so we wish to maximize the first term, which can be written as

\[ \int \log P(\phi|Z)P(\phi, Z) \, d\phi \, dZ = \int \log P(\phi|\nu)P(\phi, \nu) \, d\phi \, d\nu \]

We consider the general case of both gradient direction and magnitude, and also compare with the mutual information for just direction. By defining

\[ \tilde{s} = \sqrt{\frac{Z_1}{\sigma_1^2} + \frac{Z_2}{\sigma_2^2}} \]

and simplifying Eq. (3), the joint probability of the binding pattern and the gradient is

\[ P(\phi, Z) = \int P(\phi|Z)P(\phi, Z) \, ds \, d\phi = \sqrt{\frac{\beta}{2\pi \sigma_1^2 \sigma_2^2 I_0(\nu)}} \exp \left( \frac{-\tilde{s}^2 \cos(\phi - \tilde{\phi})}{2\sigma^2 + 4\beta} \right) \]

The joint entropy between the gradient direction and binding pattern can be approximated as follows:

\[ \int \log P(\phi, Z)P(\phi, Z) \, d\phi \, dZ \approx \log \left( \frac{2\beta}{\pi \sqrt{\pi \sigma_1^2 + 4\beta \sigma_2^2 I_0(\nu)}} \right) I_1(\nu) \]

The quality of this approximation falls when \( \beta \) increases or when \( \sigma \) decreases (data not shown). However, when \( \sigma > 8 \) and \( \beta < 8 \), the error between the approximation and the exact integral is less than 7% and quickly approaches 0 as \( \sigma \) increases or \( \beta \) decreases.

We then obtain an approximation for \( P(\tilde{Z}) \)

\[ P(\tilde{Z}) = \int_0^{2\pi} \int_0^\infty P(\phi, \tilde{Z})P(\phi, \tilde{Z}) \, d\phi \, d\tilde{Z} = \int_0^{2\pi} \int_0^\infty P(\phi)P(\phi) \, d\phi \, d\tilde{Z} \]

Defining

\[ \alpha = \arctan \frac{\kappa \sin(\nu - \tilde{\phi})}{\kappa \cos(\nu - \tilde{\phi}) + \gamma} \]

the integral can be written as

\[ \int \exp \left( \frac{\sigma^2 \nu^2 + \kappa^2 \cos(\nu - \tilde{\phi}) + \gamma^2 \cos(\nu - \tilde{\phi} + \alpha)}{2\nu^2 - 2\beta^2} \right) d\phi \]
and the identity:

$$I_0(x + a) = \sum_{k = -\infty}^{\infty} I_0(x)I_{k-4}(a) = I_0(x)I_0(a) + 2I_1(x)I_1(a)$$

for small $a$. Putting all the terms back together, we have

$$P(\overline{Z}) = \int \int P(\overline{Z} | s, \phi)P(s | \phi) ds d\phi$$

\begin{align*}
&\approx \frac{\sqrt{\beta}}{\pi \sqrt{2\sigma^2 + 4\beta}} \exp \left( \frac{\sigma^2}{2} + \frac{\beta^2}{4\sigma^2 + 8\beta} \right) \\
&\times \left[ I_0(\gamma)I_0 \left( \frac{\sigma^2}{4\sigma^2 + 8\beta} \right) + 2I_1(\gamma)I_1 \left( \frac{\sigma^2}{4\sigma^2 + 8\beta} \right) \right] \\
&\quad (18)
\end{align*}

When $\kappa < 0.4$, this approximation is accurate within 2% of the true value of $P(\overline{Z})$ for a wide range of parameters.

Transforming to polar coordinates, i.e. $d\overline{Z} = \sigma^2 d\theta d\phi$ and by numerical examination, we make the following approximations:

$$\int \log P(\overline{Z}) p(\overline{Z}) d\overline{Z} \approx \ln K + K(A_1 + A_2 + A_3 + A_4)$$

where

$$K = \frac{\sqrt{\beta}}{\pi \sqrt{2\sigma^2 + 4\beta}}$$

$$A_1 = \frac{-(\frac{1}{\sigma^2} - \frac{1}{4\beta})^2}{(\frac{1}{\sigma^2} + \frac{1}{4\beta})^{3/2}} \frac{(\kappa^2 + 8)\pi}{8}$$

$$A_2 = \frac{\sigma^2 \sqrt{2\sigma^2 + 4\beta}(3\kappa^4 + 16\kappa^2)\pi}{2\sqrt{\beta}}$$

$$A_3 = \frac{(\kappa^2 + 8)\pi}{64} \frac{\sqrt{1}}{32\sigma^2 + 4\beta}^{3/2}$$

$$A_4 = \frac{(\kappa^2 + 8)\pi}{16} \frac{1}{(\frac{1}{\sigma^2} + \frac{1}{2\sigma^2 + 4\beta})^{1/2}}$$

$$\quad (20)$$

$$\quad (21)$$

$$\quad (22)$$

$$\quad (23)$$

$$\quad (24)$$

with the constant 1.08 in $A_4$ derived from numerical approximation. For the range of parameters we tested ($\kappa < 0.4$, $\sigma \in (3, 15)$, $\beta \in (1, 8)$), the difference between the LHS and RHS of Eq. 16 is less than 3%.

Having obtained all component probabilities $P(\phi, s, \overline{Z})$, $P(\overline{Z})$, $P(\phi, s)$, we can contrast the mutual information calculated above with that in the case when the cell must estimate both the gradient steepness and gradient direction:

$$I(s, \phi, \overline{Z}) = \int \int P(\phi, s, \overline{Z}) \log \frac{P(\phi, s, \overline{Z})}{P(\phi, s)P(\overline{Z})} ds d\phi d\overline{Z}$$

As $\sigma$ increases, the cell obtains more information about the gradient from each measurement (Fig. 6). This result makes sense intuitively because $\sigma$ represents the ‘strength’ of the data, therefore the greater the $\sigma$, the more information the measurement contains. $\sigma$ is a function of the cell dimension and number of receptors, implying that the larger the cell or the more receptors the better. The joint estimation of both the gradient steepness and direction yields twice as much information as the direction estimate alone. The mutual information is also greater when the

Fig. 6. The mutual information between the binding pattern and the gradient (dashed line) or only the gradient direction (solid line). The former yields twice as much information.

Fig. 7. The numerically calculated mutual information between the binding pattern and the gradient directions for different combinations of $(\sigma_1, \sigma_2)$ for the same cell shape $(a=0.3, b=0.83)$ and 4000 receptors at ligand concentration $K_0$. There exists an optimal pair $(\sigma_1 = 7, \sigma_2 = 17.8)$ which results in the greatest mutual information.
prior distribution is broader. This is consistent with previous results in Hu et al. (2011a), where the mutual information for a fixed gradient steepness was calculated and was found to decrease as $\chi$ is higher and approaches 0 in the limit $\chi \rightarrow \infty$. Our calculation however is more general as it treats the gradient steepness also as an unknown.

The general case $\sigma_1 \neq \sigma_2$ is beyond the scope of this paper. However, by discretising $Z_1, Z_2$ to calculate the entropies, numerical calculations show that there exists an optimal combination of $(\sigma_1, \sigma_2)$ such that the mutual information $I(\phi, Z)$ is maximised. In Fig. 7, we illustrate this with one example of a cell with minor axis $a=0.3$ and major axis $b=0.83$ (to ensure the same area as a cell of diameter 1) with 4000 receptors at ligand concentration of $K_d$. Different arrangements of receptors result in different combinations of $(\sigma_1, \sigma_2)$ that can satisfy Eq. (4), and the pair that gives the greatest amount of information about the gradient is $(7, 17.8)$. This result implies that the quality of gradient sensing can be dependent on cell shape, though the lack of a formula for the general case $\sigma_1 \neq \sigma_2$ means that it is hard to address this analytically. However, in the case that the goal of the cell is to avoid bias at all cost by imposing $\sigma_1 = \sigma_2$, a highly elongated cell will be at a disadvantage compared to a circular cell of the same area because $\sigma$ is constrained by the minor axis.

5. Discussion

We have formulated gradient detection as an optimization problem, addressing in particular the question of how receptors should be arranged in order to maximize the amount of information that the sensing device can gain about the gradient. The previous works we built on (Mortimer et al., 2009, 2011; Hu et al., 2011a, 2011b; Baba et al., 2012) derived the posterior distribution of the gradient by assuming a uniform distribution of receptors on the sensing device, and a large enough number of receptors such that the Gaussian approximation holds for the magnitude and the direction of the gradient. In contrast, we use the mutual information instead of Fisher information, which makes our approach more suitable to analyse the case of a sensing device, and a large enough number of receptors such that the Gaussian approximation holds for the magnitude and the direction of the gradient. In contrast, we use the mutual information instead of Fisher information, which makes our approach more suitable to analyse the case of a circular, the cell can overcome the bias caused by its elliptical shape, which is particularly useful when the gradient is shallow. It can also maximize the information from each measurement by maximising $\sigma$, or the radius of the ‘receptor circle’. This is a general result that is independent of our choice of the prior distribution. The case $\sigma_1 \neq \sigma_2$ is beyond the scope of this paper, but it offers the enticing prospect that the cell might adjust its shape (while leaving the receptor distribution uniform) to adapt itself to exploit prevailing gradient conditions optimally.

Acknowledgements

Funding comes from an Australian Postgraduate Research Scholarship (H.N.), the Gatsby Charitable Foundation (P.D.) and Australian Research Council grant DP110101803 (G.J.G).

References